AFRL-HE-WP-TR-2004-0019

United States Air Force
Research Laboratory

AN EVALUATION OF SPEECH CONTROLS FOR AWACS WEAPONS DIRECTORS

Michael A. Vidulich
W. Todd Nelson
Robert S. Bolia
Nicole M. Guilliams
Annie B. McLaughlin
Brian P. Donnelly

HUMAN EFFECTIVENESS DIRECTORATE
COLLABORATIVE INTERFACE BRANCH
WRIGHT-PATTERSON AFB OH 45433-7022

Jeff S. Collier
Edward L. Fix
Brent Miller
Becky D. Brown
Michael R. Poole

SYTRONICS, INC.
4433 DAYTON XENIA ROAD
DAYTON OH 45432

MAY 2004

FINAL REPORT FOR THE PERIOD APRIL 2003 TO MAY 2004

Human Effectiveness Directorate
Warfighter Interface Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2004-0019

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Instruction 40-402.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

//Signed//

MARIS M. VIKMANIS
Chief, Warfighter Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE (DD-MMM-YYYY) May 2004

2. REPORT TYPE Final Report

3. DATES COVERED (From – To) April 2003 – May 2004

4. TITLE AND SUBTITLE
An Evaluation of Speech Controls for AWACS Weapons Directors

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER 62202F

5d. PROJECT NUMBER 7184

5e. TASK NUMBER 08

5f. WORKUNIT NUMBER 65

6. AUTHOR(S)
Michael A. Vidulich, *Jeff S. Collier
W. Todd Nelson, *Edward L. Fix
Robert S. Bolia, *Brent Miller
Nicole M. Guilliams, *Becky D. Brown
Annie B. McLaughlin, *Michael R. Poole
Brian P. Donnelly

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
* Sytronics, Inc.
4433 Dayton-Xenia Rd, Bldg 1
Dayton OH 45432

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSOR / MONITOR'S ACRONYM
AFRL/HECP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-WP-TR-2004-0019

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The objective of the current experiment was to evaluate speech recognition controls as a possible aid for AWACS operators. The research was carried out in the Multi-sensory Overview Large Scale Tactical Knowledge Environment (MOLTKE) facility at the Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio. The participants, trained Air Force air battle managers, were required to track a package of fighter aircraft, receive verbal target changes, and pass the verbal changes to the lead fighter aircraft in a simulated battlefield air interdiction (BAI) mission environment. The results of this research strongly supported the application of speech recognition to the TDF/PAD interface for use in AWACS applications.

15. SUBJECT TERMS
Command and Control (C2), Airborne Warning and Control System (AWACS), Speech Recognition, Interface Design, Human-Machine Interface (HMI)

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UN	UN	UN	UL	51	Michael A. Vidulich, PhD
					19b. TELEPHONE NUMBER *(Include area code)* (937) 255-8734

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE PAGE

LIST OF TABLES

EXECUTIVE SUMMARY

The objective of the current experiment was to revisit the issue of applying speech recognition controls as a possible aid for AWACS operators and determine if current speech recognition technology would be beneficial to performance in the current AWACS human-machine interface (HMI). Previous research suggested that speech recognition controls should be of value in reducing the time required to perform some tasks, potentially reducing errors, and increasing time-sharing performance in an environment requiring other manual tasks.

The research was carried out in the Multi-sensory Overview Large Scale Tactical Knowledge Environment (MOLTKE) facility at the Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio. The MOLTKE facility was a medium-fidelity simulated air battle management environment comprising six PC-based operator workstations running Prototype AWACS Display (PAD). The participants, trained Air Force air battle managers, were required to track a package of fighter aircraft, receive verbal target changes, and pass the verbal changes to the lead fighter aircraft. These actions were carried out in a simulated battlefield air interdiction (BAI) mission environment.

The results of this research strongly supported the application of speech recognition to the TDF/PAD interface for use in AWACS applications. The use of speech as a control modality allowed operators to perform some tasks, such as system set-up or the aircraft-target repairings, more quickly than they could using the baseline manual controls. There was also evidence that the availability of speech controls allowed for more efficient time-sharing performance.

The operators that participated in the study were very favorable toward the use of speech controls. This was demonstrated not only by their subjective workload ratings and questionnaire responses, but most directly by their behavior in preference trials. Given a choice of using or not using speech for the speech enabled tasks, the operators overwhelmingly chose to do so.

INTRODUCTION

Air Battle Management (ABM) has become a vital component of the U.S. Air Force's warfighting capability (Vidulich, Bolia, & Nelson, in press). An essential part of the U.S. Air Force's ABM is the Airborne Warning and Control System (AWACS) (Armistead, 2002; Hirst, 1983; Isby, 1997). Boyne (2003) summarized the impact of the AWACS aircraft:

> Since its debut, the AWACS has performed brilliantly domestically and in conflicts around the world, including those in Panama, Grenada, Haiti, Kuwait, Kosovo, Serbia, Afghanistan, and Iraq. In the 1991 Gulf War, the all-seeing eye of the AWACS completely encompassed not only the country of Kuwait but the entire area of responsibility (AOR) as well, detecting and noting where every aircraft flew every moment of the day. The AWACS provided information to guide more than 120,000 sorties by coalition forces, flying more than 5,000 hours on 400 missions to do so. It took part in thirty-eight of the forty aerial victories gained by coalition air forces. (p. 67)

Given the important central role of the AWACS aircraft and the level of air battle activity an AWACS aircraft may sometimes be responsible for monitoring and controlling during a large campaign, the need to optimize the AWACS operator's human-machine interface (HMI) is obvious. The current paper will review an evaluation of speech controls as a potential addition to the AWACS interface. As background to that evaluation, the role of the AWACS as a Command and Control (C2) asset and the current status of speech recognition as a potential interface component will first be reviewed.

Command and Control (C2) Task Domain

C2 activities have long been recognized as a vital part of military operations. The Department of Defense defines C2 in Joint Publication 1-02 as:

> The exercise of authority and direction by a properly designated commander over assigned forces in the accomplishment of the mission. Command and control functions are performed through an arrangement of personnel, equipment, communications, facilities, and procedures which are employed by a commander in planning, directing, coordinating, and controlling forces and operations in the accomplishment of the mission (JP 1-02, p.100).

More specifically, "command" is the authority vested in individuals in order to clearly delineate responsibility and foster unity of effort through the direction, coordination, and supervision of military forces. Activities involved with "control", include collecting, processing, displaying, storing, and disseminating information for the use during planning, preparing for, executing, and assessing operations (Sweeny, 2002). Historically, C2 is not a new concept. From Alexander the Great's shouted battlefield commands to recent "information age engagement," those who have mastered the techniques and applications of C2 have most often prevailed in combat (Sweeny, 2002). It has also been seen in conflicts such as the Battle of Britain, where controllers in a Group and Sector Operations Room analyzed all available information, and provided information to the flyers on topics such as intercept points, weather, and enemy actions (Air Ministry Account, 1940).

The C2 process is often carried out on many different levels, and varies depending on the situation. The Joint Operations Center, command post, or Air Operations Centers are centers for C2 during a large-scale battle or war. On a lower echelon of command, the Airborne Warning and Control System (AWACS) is used to oversee air battle spaces. The AWACS provides highly mobile, flexible, and survivable wide-area surveillance and C2 functions while en route to, or upon arrival at its destination (Armistead, 2002). It operates in conjunction with other C2 systems to integrate airborne elements of command, air surveillance, air support coordination, airspace management, and control. Further enhancing the operational stature of the AWACS, is its ability to conduct all-weather surveillance above various types of terrain (Armistead, 2002). The officers that fly the AWACS are generally divided into the flight deck personnel (pilots and navigators) and the mission crew consisting of technicians, surveillance personnel, and Senior or Weapons Directors (WDs).

The WDs are responsible for the C2 of aircraft within a particular airborne battle space and at times, assist commanders in making battle space decisions. As stated by Klinger, Andriole, Militello, Adelman, & Klein (1993), WDs can be compared to Air Traffic Controllers in the sky, with some important differences: Air Traffic Controllers are never being fired upon, never need to monitor enemy aircraft, are not traveling with the aircraft they are controlling, and they need not to worry about the rules of engagement. Another important difference between Air Traffic Controllers and WDs is that WDs are primarily concerned with directing aircraft toward each other, such as air refueling, rather than trying to keep aircraft apart (Fahey, Rowe, Dunlap, & deBoom, 2001).

C2 in the WD's domain can also be called air battle management (ABM), which is a complex and demanding task that involves directing the implementation of the air tasking order (ATO) and controlling the execution of the associated air-to-air and air-to-ground operations. The ATO is a method used to task and disseminate to components, subordinate units, and C2 agencies projected sorties, capabilities and/or forces to targets and specific missions (JP 1-02, p. 29). It normally provides specific instructions to include call signs, targets, controlling agencies, etc., as well as general instructions. Along with the ATO, mission objectives might include patrolling an inactive airspace, reconnaissance, launching defensive counter air strike in response to an enemy attack, controlling an offensive strike package into enemy territory, as well as non-combat missions such as search and rescue. ABM involves the painstaking monitoring and manipulation of data-saturated situation displays, comprising maps overlaid with landmarks, geographical features, and moving tracks representing the air and ground assets of coalition and enemy forces. It also involves the monitoring and manual control of up to eight communication channels, including radios and intercoms. Performing these tasks requires tracking an enormous amount of information, such as the position, heading, altitude, and speed of both friendly and hostile aircraft, and the fuel and armament status of friendly aircraft (Fahey et al., 2001). The majority of these tasks are performed at the WD workstation, which includes panels of toggle switches, knobs, and dials, a trackball, keyboard, and numerous reconfigurable push buttons. Although functional, such workstations are manually-intensive and may require extensive training and introduce inefficient control actions. During periods of low to moderate air traffic, an ABM can comfortably manage the tasks. However, during periods of heavier air traffic, operators are very likely to reach unacceptable levels of workload, which may negatively impact performance efficiency and mission effectiveness (Nelson et al., 2003).

Speech Recognition

Salisbury and Chilcote (1989) suggested that speech recognition might be an effective addition to the human-machine interface used by AWACS operators. Some basic functionality was demonstrated and evaluated in a rapid prototyping facility. At that time, the results of the evaluation were mixed. User assessments favored the intuitive nature of speech commands, but the experimenters noted that "Robustness, however, remains a problem, especially for voice input" (Salisbury & Chilcote, 1989, p. 55).

Technological Progress. In the years since the Salisbury and Chilcote study, speech technology has changed considerably. Automatic speech recognition is a now reasonably mature control technology that has been under development for almost three decades (Nelson et al., 2003). Early systems had small vocabularies, required speaker training, and demanded unnatural pauses between words, but recent advances in two of the basic component areas – signal processing and control algorithms – that have permitted the development of commercially available speech-based control systems capable of recognizing continuous, speaker-independent speech input (McMillan, Eggleston, & Anderson, 1997). For example, automatic speech recognition technology has been successfully implemented in telephone call handling systems, telephone dialing technology, speech-based control of numerous appliances and devices for physically-disabled users, and telephone-based banking and credit card systems. Even more advanced speech systems are now helping customers look up names, answer questions about mortgages, trade stocks and options nationwide, book flight and hotel reservations, and browse the web (Judge & Browder, 1998). Some speech recognition systems can now transcribe unbroken speech with up to 99 percent accuracy (Roush, 2003).

In addition to the civilian commercial uses of speech recognition, speech-based control has also been tested and applied in military applications. For example, speech-controls have been incorporated into the multi-function display of an experimental F-16 fighter aircraft (McMillan, Eggleston, Anderson, 1997), a simulated unmanned aerial vehicle operator (UAV) workstation (Draper, Calhoun, Ruff, Williamson, & Barry, 2003), and is also being considered for incorporation into the next-generation Joint Strike Fighter (Proctor, 2000). Speech-based control was also used in a Theater Air Planning (TAP) module for data entry application (Williamson & Barry, 2000). In the TAP application, speech-based input was found to be robust to differences in voice styles and dialects. The overall recognition rates observed in the TAP and UAV applications both exceeded 97% (Williamson, Barry, and Draper, in press). Comparable recognition accuracy has also been demonstrated for speech generated in a noisy aircraft environment, even under positive sustained acceleration (Williamson, 1999).

Speech Recognition Benefits. Perhaps the most notable advantage of speech recognition is that for some tasks it provides time-savings compared to the comparable manual interfaces. For example, in the Theater Air Planning application discussed above, a 10.6% time savings was observed; in the UAV application, data entry times were reduced by approximately 40% (Williamson et al., in press). In the UAV evaluation, speech recognition also dramatically increased the number of successful task completions.

Designing an interface that uses both speech and manual controls has also been viewed as a way to increase time-sharing performance in complex task environments (Vidulich, 1988; Vidulich & Bortolussi, 1988a; Wickens, Sandry, & Vidulich, 1983; Wickens, Vidulich, & Sandry-Garza, 1984). The most common explanation for the *time-sharing benefit* has been the multiple resource model (Wickens, 1980), by which the control of manual and speech responses are attributed to separate attentional resource pools. So, mixed manual and speech responses are expected to compete with each other less than multiple responses of either single type. In addition to laboratory demonstrations of this advantage (e.g., Vidulich, 1988; Vidulich & Wickens, 1981; Wickens, Vidulich, Sandry, & Schiflett, 1981), there have been several demonstrations in simulations of complex real-world tasks. For example, Wickens, Sandy, and Vidulich (1983) demonstrated in a F/A-18 mock-up simulator that a speech response for a secondary task was time-shared better with flight control than a manual response. Also, studies of time-sharing multiple tasks during a simulated helicopter hovering task demonstrated that manual flight control inputs increased and helicopter movement decreased if the additional tasks required speech inputs rather than manual inputs with the free hand (Bortolussi & Vidulich, 1989; Vidulich & Bortolussi, 1988b). Although the time-sharing performance benefits observed in the helicopter simulations were reliable, subjective workload ratings suggested that speech control was associated with greater effort due to the need to speak precisely (Bortolussi & Vidulich, 1991; Vidulich & Bortolussi, 1988a).

Experimental Objectives

The objective of the current experiment was to revisit the issue of applying speech recognition controls as a possible aid for AWACS operators and determine if current speech recognition technology would be beneficial to performance in the current AWACS human-machine interface (HMI). Previous research suggested that speech recognition controls should be of value in reducing the time required to perform some tasks, potentially reducing errors, and increasing time-sharing performance in an environment requiring other manual tasks. Although earlier performance benefits were sometimes linked to an increase in mental workload, the mental workload increase may have been the result of the brittle recognition performance available at that time.

However, the AWACS environment, in addition to requiring a great deal of manual control (e.g., button pressing, mouse movements), is also communications intense. Consequently, the benefits or disadvantages of speech for an AWACS environment must be evaluated in an environment with realistic communication tasks.

METHOD

Participants

Twelve active-duty officers (11 male, 1 female) from the United States Air Force assigned to the 552nd Air Control Wing at Tinker Air Force Base, Oklahoma served as the participants in this evaluation. Eleven of the participants reported being right-handed and one reported being left-handed. The average age of all participants was 27.8 years old. All participants had the Air Force Specialty Code 13B, Air Battle Manager. Eight participants were classified as Mission Ready (MR) to be deployed on the AWACS. Of these eight, four had participated in flag exercises in the previous three years and seven had been deployed within that time-period. The average MR hours flown by the eight MR participants was 616 hours, with a range from 0 hours to 1500 hours. Three of the remaining participants were classified as Basic Qualified (BQ). The BQ participants were qualified to fly AWACS training missions within the United States but were not to be deployed. The remaining participant, although trained as a 13B Air Battle Manger, was rated as Unqualified (UQ) on the AWACS aircraft.

Apparatus

The experiment was conducted in the Multi-sensory Overview Large-scale Tactical Knowledge Environment (MOLTKE) facility. The MOLTKE facility is a medium-fidelity simulation of an AWACS environment with six workstations arranged in two rows of three facing each other, similar to a console arrangement on the AWACS E-3 aircraft. Each workstation consisted of two 900 MHz computers, one 19" flat panel-display, keyboard, mouse, programmable keypad, audio control panel, and two footswitches. The workstation computers used the Microsoft Windows 2000 operating system and the Solipsys Tactical Display Format – Prototype AWACS Display (TDF/PAD) software (Conn, 2003). The TDF/PAD will be the next generation AWACS platform, and was viewed by the C2 community as the "best technology for achieving an intuitive common look-and-feel for both ground and airborne platforms" (Conn, 2003). In combination with the TDF/PAD, a commercial-off-the-shelf speech system, Nuance 8.0, was used.

The six workstations were connected to an experimenter's control station. The experimenter's control stations was composed of several computers that ran the AuSIM spatial audio system, A/D and D/A converters, the MSCT software, and the experimental control software.

This present experiment made use of five workstations. The participant performing the role of a Weapons Director (WD) occupied Station 1. Three other stations were occupied by the experimental White Team members. Station 3 was occupied by an experimenter playing the role of Observer, whose job it was to classify events using the programmable keypad. Station 4 was occupied by an experimenter playing the role of Senior Director (SD). Station 6 (Strike) was occupied by an experimenter playing the role of all of the pilots in the various strike packages controlled by the WD. The participant, SD and the observer workstations all ran the PAD interface software to observe the air battles being conducted in the various experimental scenarios. The Strike workstation ran the ModSAF simulation software that actually provided the battle data for the PAD software to display. Strike updated the ModSAF software to reflect the control received from WD.

Several software packages were resident within the MOLTKE system to conduct the evaluation. These included: (1) Solipsys TDF 3.7 that supported the PAD interface; (2) ModSAF, which generated the constructive battle simulation according to Distributed Interactive Simulation (DIS) PDU protocols; and (3) MSCT, which linked the ModSAF PDUs to the TDF/PAD interface. The DIS PDU data traffic was collected by hlaResults software. As an additional backup data collection, the CG2 DIS Logger was also employed.

The participant's PAD display with a video feed of the participant in the upper-right corner and the intercom audio communications were recorded on DVD for post experimental review.

Mission Task

Each trial's mission consisted of four phases: set-up, ingress, retargeting, and egress. In the set-up phase, the participant configured the WD station to prepare for the mission. This consisted of sorting the ATO list, marking aircraft with their respective controllers, and marking the initial bulls-eye for the trial. This phase was completed without any active ModSAF simulation running. The set-up phase started when the participant, prompted by an experimenter, pressed a button to indicate the beginning of the phase and ended when the participant pressed the start button to start the simulation. The ingress phase consisted of the four strike packages moving towards their pre-assigned targets. The participant would monitor their movements, respond to any requests for information from the strike packages (i.e., "picture calls"), inform Strike if any enemy aircraft were in a position to attack a strike package (i.e., "threat calls"), and set-up pairing lines on the display to connect the four strike packages to their assigned targets. At a designated point (unknown to the participant), the SD would inform the participant that the targets for the four strike packages had changed. The SD would provide four sets of "nine-line" data to the participant to identify the new targets. The participant wrote down the new information and conveyed it to the appropriate strike packages. The participant also updated the PAD interface to show the new strike package and target assignments with pairing lines on the PAD interface. Following the transmission of the four nine-lines to Strike, the egress phase began. At this point, the participant continued to monitor for any threats to the strike packages as they attacked their new targets. After the last strike package attacked its target and started back, Strike called "knock-it-off" and pressed a button to terminate the trial.

Loading Tasks

In addition to the main mission task, the participant also performed two types of loading tasks: the coordinate response measure (CRM) task and three probe tasks.

Coordinate Response Measure (CRM) task. The CRM task (refer to Bolia, Nelson, Ericson, & Simpson, 2000, for a description) provided a measure of communications intelligibility. The CRM task has been successfully used to assess communications in-flight on an AWACS aircraft (Bolia, 2003). In brief, the CRM consists of short phrases comprising a call sign followed by a color-number combination (e.g., "Ready Baron, go to Blue Five now."). In this experiment the CRM ran continuously throughout the ingress, retargeting, and egress phases. Participants were instructed to listen for a specific call sign ("Baron") and, if detected, to enter the color-number combination contained in the phrase. Detection responses were issued by pressing the appropriate color-number combination on the CRM response pad, a 4×8 matrix of colored digits.

Probe Tasks. Three probe tasks were used to give the participant multiple tasks to handle throughout the trial. The three tasks were: ATO Query, Bearing and Range Determination, and Switch Bulls-eye. All three probe tasks could be performed either with manual response or by speech command. The ATO Query probe task started when the SD, at a pre-designated time, asked the participant for a piece of information in the ATO. The participant used either manual or speech commands to find the required information and replied to SD over the intercom. The Bearing and Range task started when the SD, at a pre-designated time, asked the participant for the bearing and range from one of the strike packages to a target. The bearing and range probe task only occurred following the retargeting of the strike packages. The bearing and range probe task always required some manual inputs from the participant to accomplish, but in the speech control modality trials there were also speech-enabled components of the task. The participant always designated a bulls-eye location as part of the set-up phase. At some point in each trial the SD would command the participant to change the bulls-eye location. As with the other probe tasks, the Switch Bulls-eye probe task could be performed either with manual commands or with speech commands. For all three probe tasks, the three members of the white team marked the beginning and end of the task by use of their programmable keypads.

Main Independent Variable

The main independent variable in the evaluation was control modality. Numerous tasks during the simulation could be performed either by using manual inputs (keyboard or mouse) or by speech commands. On designated trials the participants would be instructed to respond to all tasks using manual inputs (All-Manual Trials) or to use speech on any speech-enabled task (Speech Trials). There was also one trial – the *Preference Trial* – for which the participants were instructed to use either control modality for any task.

Experimental Design

The experiment was conducted within-subject. For the main experiment, each participant performed three blocks of two trials. One trial was performed with all-manual controls, the other utilized speech controls where possible. For each participant, one control modality condition was used first in each block. For odd-numbered participants, each block started with a speech trial, whereas for even-numbered participants each block began with an all-manual trial.

Six different scenarios (i.e., simulated missions with specific terrain, preplanned strike package routes and targets, retargeting nine-line sets, and probe tasks) were used in the main experiment and were distributed randomly to different block and control modality conditions across participants.

The preference trial was performed only once by each participant. It was performed last, following all six trials of the main experiment. The same scenario was used for all participants in the preference trials.

Procedure

Each participant started with a period of training to familiarize them with the PAD interface, the speech commands, and the experimental task. This was followed by the six trials of the main experiment and the single preference trial.

Training. The training period typically required about four to six hours. The basic steps followed during training were: an introduction to the PAD, introduction to the battlefield air interdiction (BAI) scenario, training to complete tasks manually, training to complete the tasks using speech, and several practice trials.

Experimental Trials. Each trial started with an experimenter informing the participant of the control modality condition for the upcoming trial and telling them to start when ready. The participant proceeded by pressing the trial start button and performing the set-up phase. As soon as the set-up was completed, the participant pressed another button to end the set-up phase and begin the ModSAF scenario. Each scenario progressed through the ingress, retargeting, and egress phases and was terminated by Strike's knock-it-off call. At the end of each trial, the participant filled out a NASA-TLX form. The NASA-TLX form included independent ratings of all four trial phases. Following completion of the six main experimental trials and the preference trial, the participants filled-out debriefing questionnaires and were debriefed about the experiment.

Dependent Measures

The simulation computers collected all DIS PDUs generated during every trial. These data included the various interface inputs to the TDF/PAD by the participant and the human factors inputs by the white team. Subjective data (i.e., NASA-TLX ratings and the participant's preferences in the preference trial were assessed post hoc by review of video recordings.

RESULTS

There were nine major categories of data collected in the experiment: trial duration, mission performance efficiency, NASA-TLX mental workload ratings, probe task performance, communications behavior, intercom speech intelligibility, TDF/PAD interface behavior, preference trial selections, and debriefing questionnaire responses. Analyses of the first seven categories will focus on the six main experimental trials performed by each participant. The preference trial and questionnaire response data will be considered independently.

Trial Duration

Ideally, the analysis of trial duration would include a breakdown of duration by mission phases. Unfortunately, due to unreliability in the marking of the transition from the retargeting phase to the egress phase in the raw data, the analysis by phase was not possible. Nevertheless, analysis of the overall trial duration was undertaken to identify whether either control modality or block affected it and to establish a trial length baseline for later use. The overall duration of all trials were analyzed with a Control Modality x Block analysis of variance (ANOVA). No significant effect involving either variable was detected. The overall mean duration of experimental trials was 1076 s (or 17 m 56 s).

Mission Performance Efficiency

Mission performance efficiency refers to the participants' speed in performing the basic tasks that they would normally do as part of a real mission. Four such measurements were collected on each experimental trial: (1) *Set-up Phase Duration* – the time required to complete the configuration of the interface in set-up phase of the trial; (2) *Initial Pairings Time* – the time required to accomplish the four initial pairings of strike packages and targets in the ingress phase; (3) *Nine-Line Transmission Time* – the time required to receive the four nine-line transmissions from SD and to transmit them to Strike; and (4) *Strike Packages Repairing Times* – the time required to update the pairings information to correspond to the new missions.

<u>*Set-up Phase Duration*</u>. The times between the participants' two button presses indicating the starting and the completion of the set-up phase were collected during the main experimental trials. These data were analyzed with a 2 (Control Modality) × 3 (Block) repeated measures ANOVA. As illustrated in Figure 1, the duration of the set-up phase was significantly shorter in the speech condition (80 s) than in the all-manual condition (102 s), $F(1, 11) = 6.968, p = .023$. There was also a significant effect of block due to the duration of the third block being shorter than the first two blocks (97 s, 97 s, and 80 s, across the three blocks), $F(2, 22) = 4.936, p = .017$.

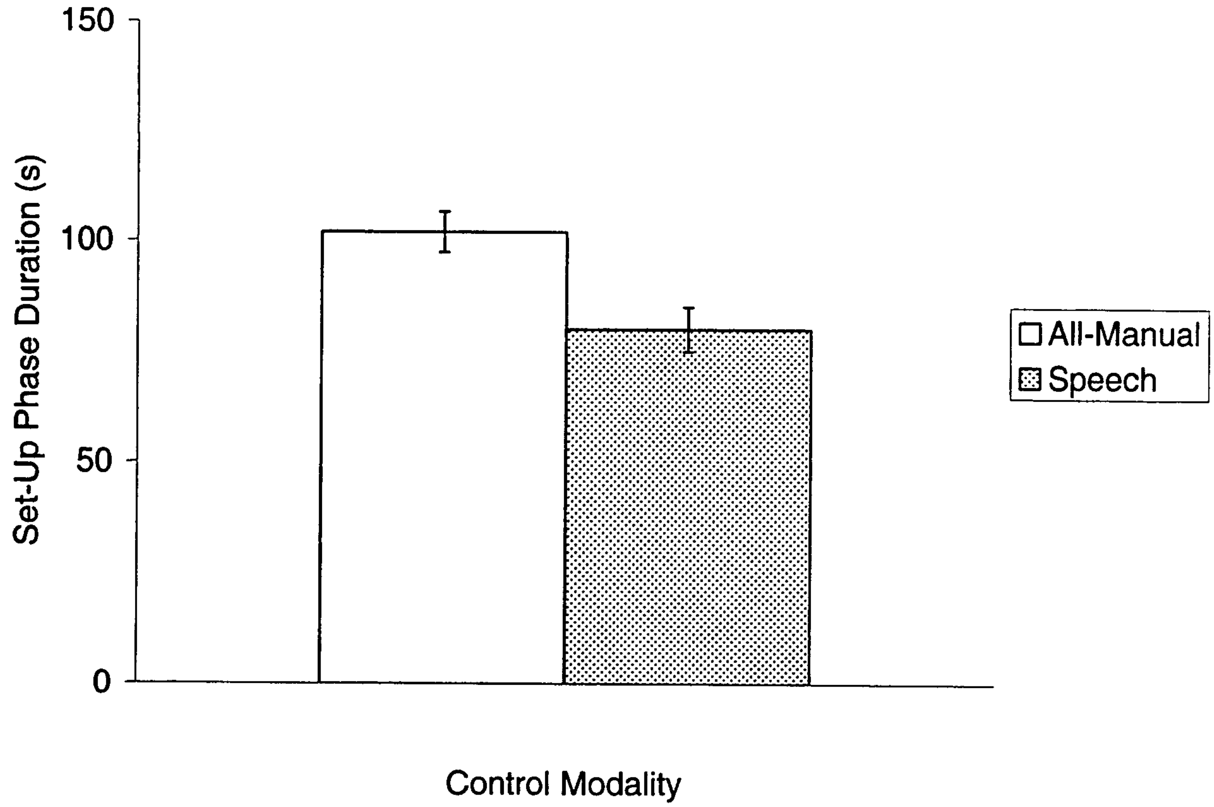

Figure 1. The mean durations and standard errors of the set-up phase as a function of control modality.

Initial Pairings Time. There were a considerable number of failures to perform the four expected initial pairings during the ingress phase. Out of a total of 72 trials, the participants failed in 30 (42%) to achieve a minimum of four pairings in the ingress phase and created excess pairings during 3 trials (4%). The participants varied considerably in their accomplishment of initial pairings. One participant never performed any initial pairing in any of the six main experimental trials, another participant only executed initial pairings in one trial. On the other hand, three participants performed the recommended four pairings on all six of their trials. The initial pairings timing for every successful initial pairings trial (i.e., a minimum of four pairings in place at the end of the ingress phase) were calculated by subtracting the time of the first pairing event in the ingress phase from the time of the final pairing event in the ingress phase. The resulting timing data for all participants that accomplished at least one successful trial in each control modality condition were analyzed by a one-way (Control Modality) ANOVA. The initial pairings were always performed manually regardless of the trial's control modality, so no difference was expected. The results confirmed the expectation (all-manual = 90.3 s, speech = 89.7, $p = .950$).

Nine-Line Transmission Time. The times between the SD's first transmission of a new mission nine-line to the participant and the participant's last nine-line transmission time to Strike were collected on all six main experimental trials. These data were analyzed with a 2 (Control Modality) × 3 (Block) repeated measures ANOVA. As illustrated in Figure 2, the time required for the nine-line transmission task was significantly shorter in the speech condition (361 s) than in the all-manual condition (407 s), $F(1, 11) = 5.501$, $p = .039$. Neither the main effect of Block nor the Control Modality × Block interaction reached significance, $p > .05$.

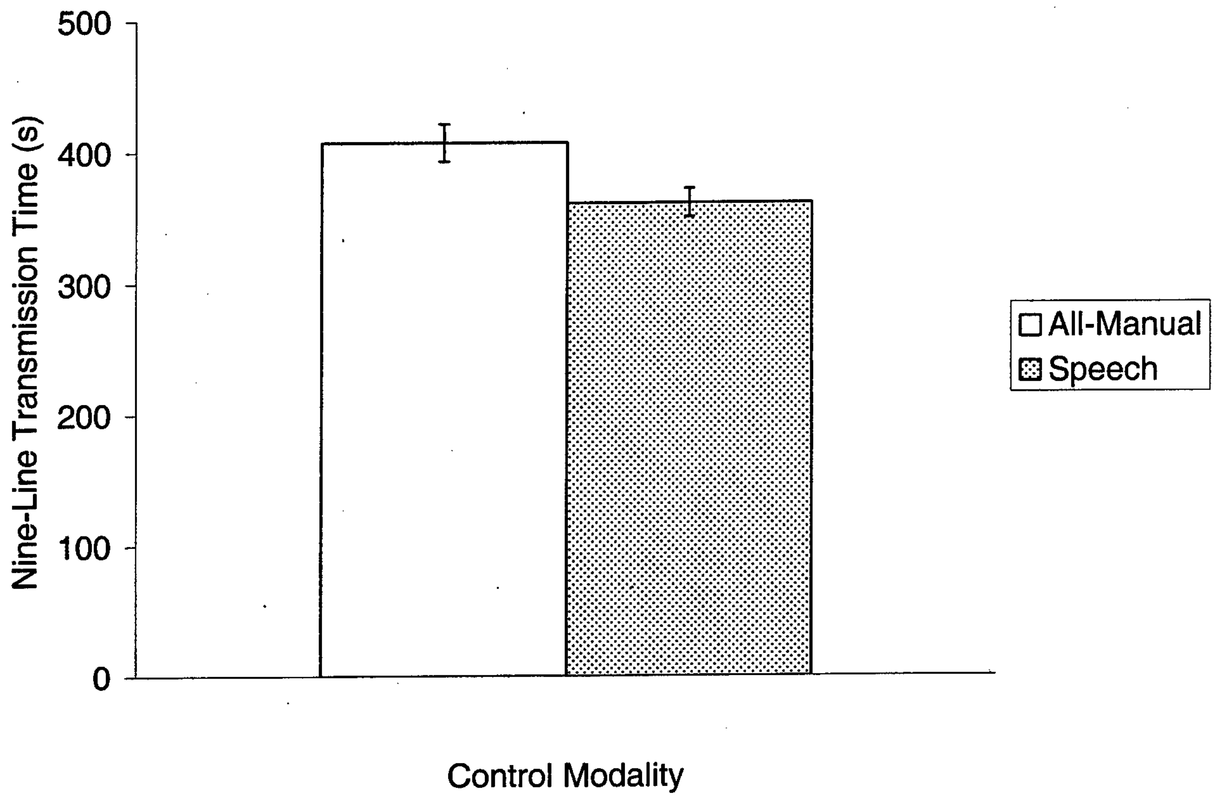

Figure 2. The mean durations and standard errors of nine-line transmission during the Retargeting Phase as a function of control modality.

Strike Packages Repairing Time. Following the ingress phase, the participant would receive instructions for four strike package target changes. Along with transmitting these changes to Strike, the participant was expected to update the pairings on their situation display. The participants were much more successful in generating the four new pairings following the ingress phase than they were in performing the initial pairings. On only three trials (4%) did the participants fail to generate four new pairings. During 15 trials (21%), however, the participants generated an excess of pairings following the end of the ingress phase. In performing the repairings, the participant could, and were instructed to, use speech commands during the speech control modality trials. The repairing times were analyzed by a one-way (Control Modality) ANOVA. As illustrated in Figure 3, the repairings were accomplished 77 s faster in the speech condition than in the all-manual condition, $F(1, 11) = 11.026$, $p = .007$.

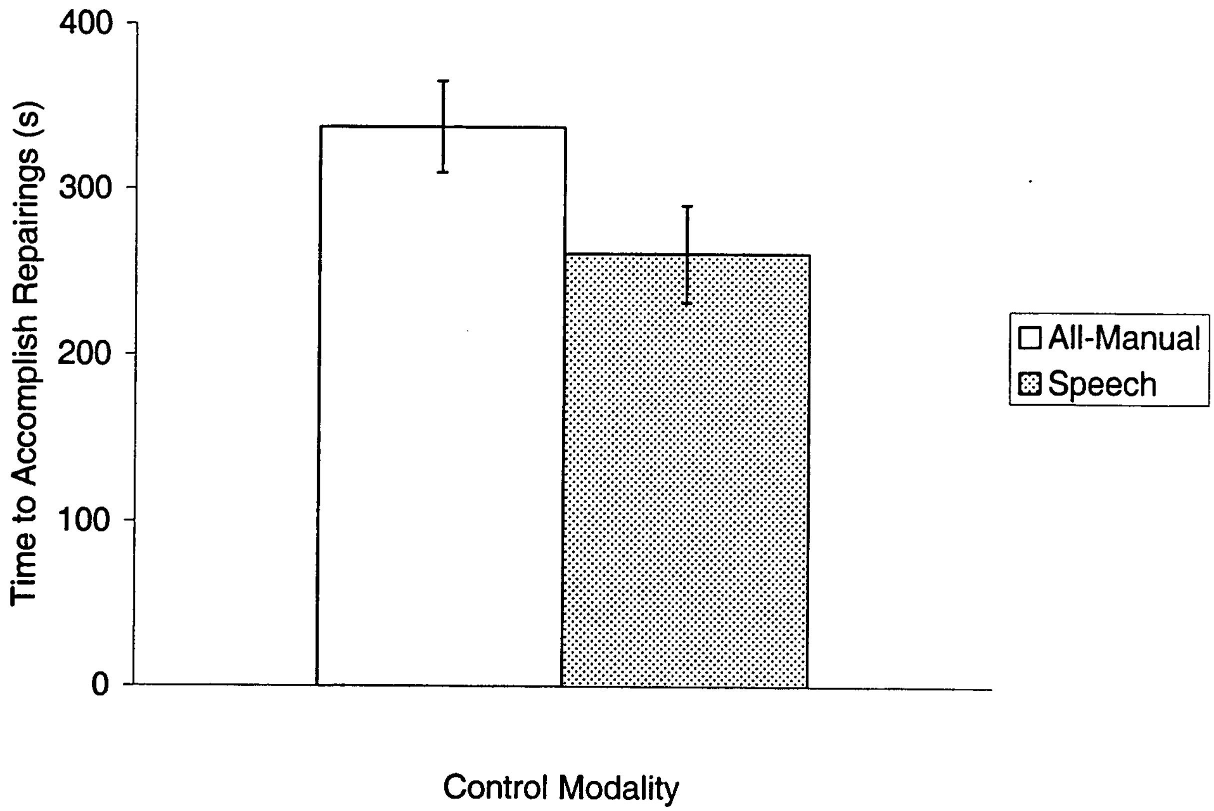

Figure 3. The mean times and standard errors to accomplish the repairings of the four strike packages as a function of control modality.

NASA-TLX Mental Workload Ratings

The NASA-TLX ratings were analyzed with a 2 (Control Modality) × 3 (Block) × 4 (Phase) repeated measures ANOVA. There was a significant main effect of Block, $F(2, 22) = 9.370$, $p = .001$ (Means: Block 1 = 44, Block 2 = 39, Block 3 = 31). There was also a main effect of Phase, $F(3, 33) = 17.164$, $p < .001$ (Means: set-up = 34, ingress = 43, retargeting = 47, egress 28).

There was a significant Control Modality x Phase Interaction, $F(3, 33) = 2.987$, $p = .045$. As illustrated in Figure 4, the speech condition was rated somewhat lower in all four phases, but especially so in the retargeting phase.

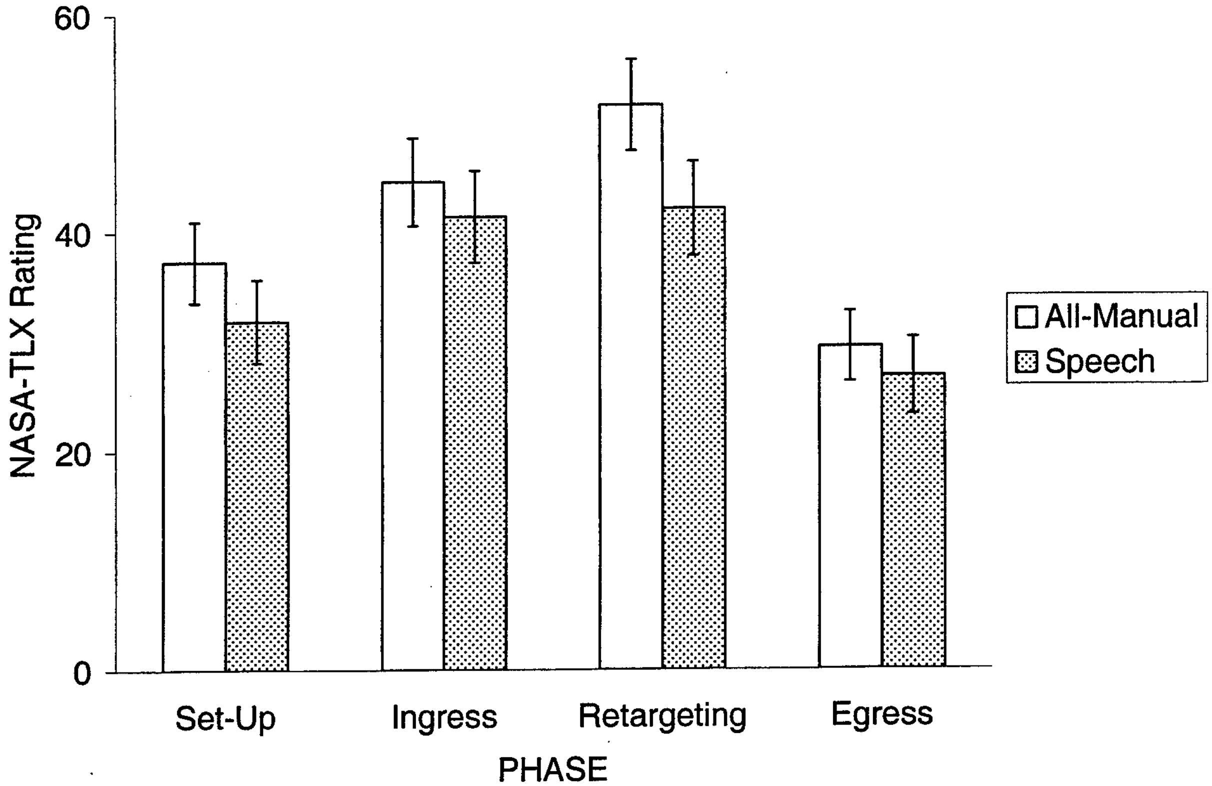

Figure 4. The mean NASA-TLX ratings of mental workload and standard errors as a function of phase and control modality.

As a diagnostic aid, the raw scores on each of the six NASA-TLX scales were analyzed with individual 2 (Control Modality) × 3 (Block) × 4 (Phase) repeated measures ANOVAs.

<u>Mental Demand</u>. There were statistically-significant main effects of Block ($F(2, 22) = 8.890, p = .001$) and Phase ($F(3, 33) = 14.872, p < .001$) detected in the mental demand ratings. The average mental demand rating decreased across the three blocks of the experiment (43, 37, and 31, respectively). The highest-rated phase was retargeting (46), followed by ingress (42), set-up (33), and egress (26). No other main effect or interaction reached significance ($p > .05$).

<u>Physical Demand</u>. There were statistically significant main effects of Control Modality ($F(1, 11) = 9.871, p = .009$), Block ($F(2, 22) = 4.895, p = .017$), and Phase ($F(3, 33) = 9.361, p < .001$) detected in the physical demand ratings. The all-manual control modality trials were rated as more physically demanding (29) than the speech control modality trials (19). The average physical demand rating decreased across the three blocks of the experiment (26, 25, and 20, respectively). The highest-rated phase was retargeting (31), followed by ingress (25), set-up (22), and egress (17). A significant Control Modality x

Phase interaction was also detected, $F(3, 33) = 5.222, p < .005$. The interaction is displayed in Figure 5. No other interaction reached significance ($p > .05$)

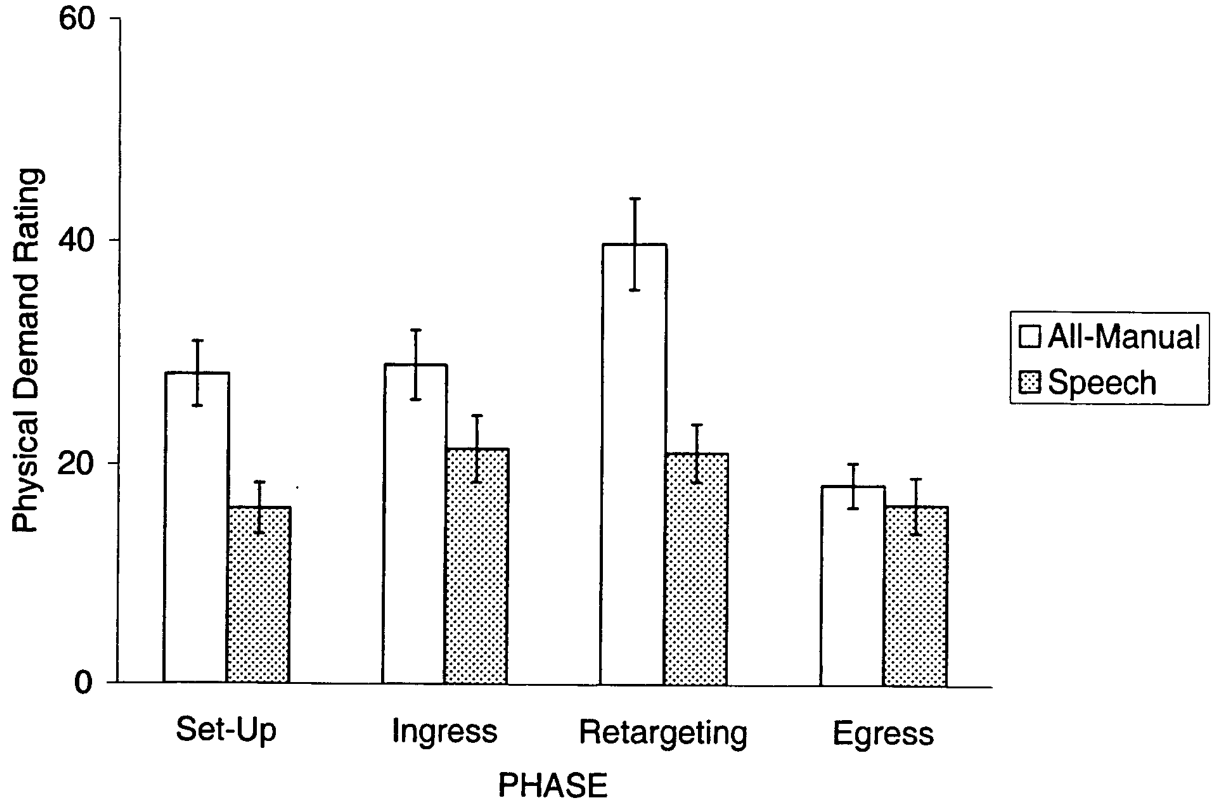

Figure 5. The mean Physical Demand ratings and standard errors as a function of phase and control modality.

<u>Temporal Demand</u>. There were statistically-significant main effects of Block ($F(2, 22) = 11.387, p < .001$) and Phase ($F(3, 33) = 20.071, p < .001$) detected in the temporal demand ratings. The average temporal demand rating decreased across the three blocks of the experiment (44, 39, and 31, respectively). The highest-rated phase was retargeting (50), followed by ingress (44), set-up (36), and egress (21). No other main effect or interaction reached significance ($p > .05$).

<u>Performance</u>. There were statistically-significant main effects of Block ($F(2, 22) = 8.994, p = .001$) and Phase ($F(3, 33) = 8.930, p < .001$) detected in the performance ratings. The average performance rating decreased across the three blocks of the experiment (43, 38, and 24, respectively). The highest-rated phase was retargeting (43), followed by ingress (40), set-up (32), and egress (31). No other main effect or interaction reached significance ($p > .05$).

Effort. There were statistically significant main effects of Block ($F(2, 22) = 6.819, p = .005$), and Phase ($F(3, 33) = 12.004, p < .001$) detected in the effort ratings. The average effort rating decreased across the three blocks of the experiment (45, 43, and 34, respectively). The highest-rated phase was retargeting (48), followed by ingress (45), set-up (40), and egress (31). A significant Control Modality x Phase interaction was also detected, $F(3, 33) = 3.681, p < .022$. The interaction is displayed in Figure 6. No other main effect or interaction reached significance ($p > .05$)

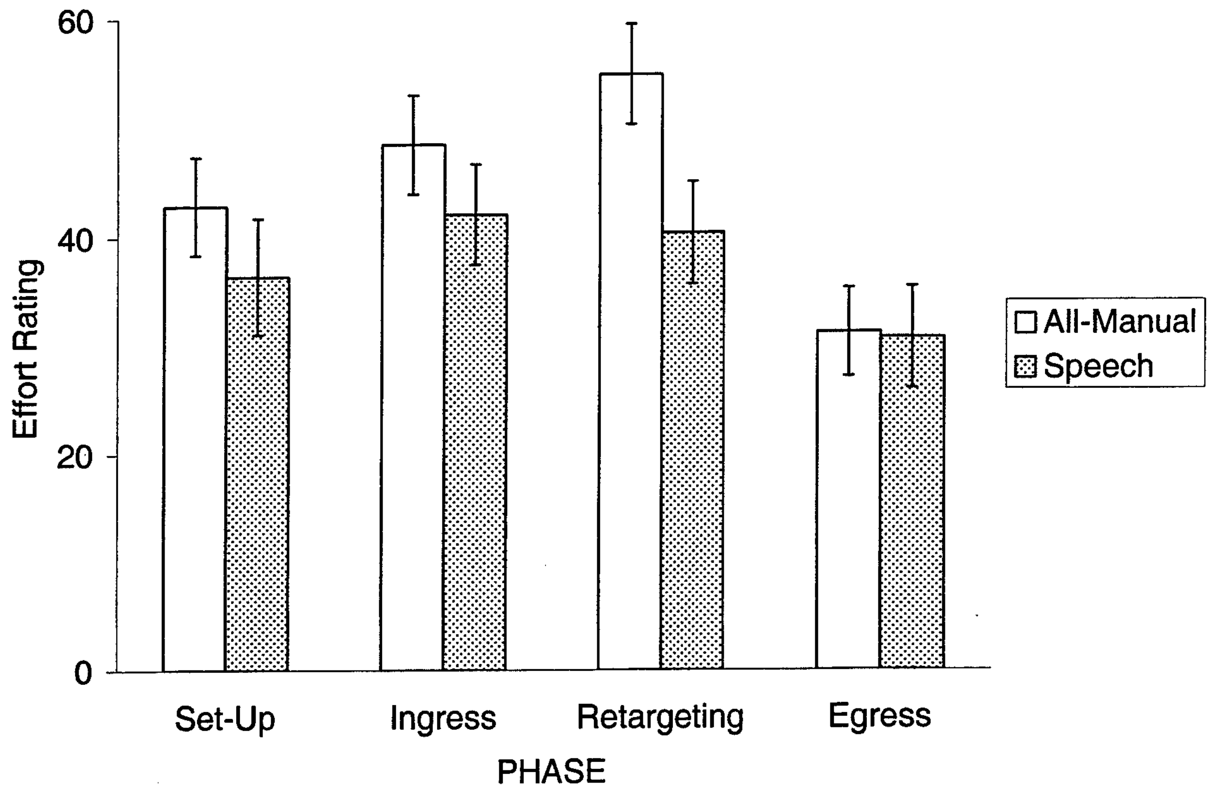

Figure 6. The mean Effort ratings and standard errors as a function of phase and control modality.

Frustration. There were statistically-significant main effects of Block ($F(2, 22) = 5.527, p = .011$) and Phase ($F(3, 33) = 11.860, p < .001$) detected in the frustration ratings. The average frustration rating decreased across the three blocks of the experiment (44, 40, and 31, respectively). The highest-rated phase was retargeting (48), followed by ingress (44), set-up (33), and egress (28). No other main effect or interaction reached significance ($p > .05$).

Overall, all six NASA-TLX dimensions detected a reduction in subjective demand over the three blocks of the experiment and the same ordering of demands across the four phases of the scenarios. The only main effect of control modality was in the beneficial effect of the speech control modality detected in the physical demand dimensions. Two dimensions, physical demand and effort, detected the Control Modality x Phase interaction that was significant in the overall NASA-TLX ratings. The pattern of means and standard errors involved in the two individual scale interactions (Figures 5 and 6) were much the same as that seen in the overall NASA-TLX interaction (Figure 4), except for more separation between the control modalities in the set-up phase in the physical demand ratings than for the overall NASA-TLX or the effort ratings.

Probe Task Performance

The error rates for the ATO question probe task and the switch bulls-eye probe task were both 2.7% over all trials. The error rate for the bearing and range probe task was somewhat higher at 9.7%. The response times for all correct trials were culled of outliers with reaction times exceeding 30 s. The remaining data were analyzed by a one-way (Control Modality) ANOVA. As illustrated in Figure 7, the only significant effect was detected in the ATO-question ANOVA. Participants were significantly faster at answering the ATO question probes in the all-manual condition (6.2 s) than in the speech condition (11.0 s), $F(1, 11) = 57.642, p < .001$.

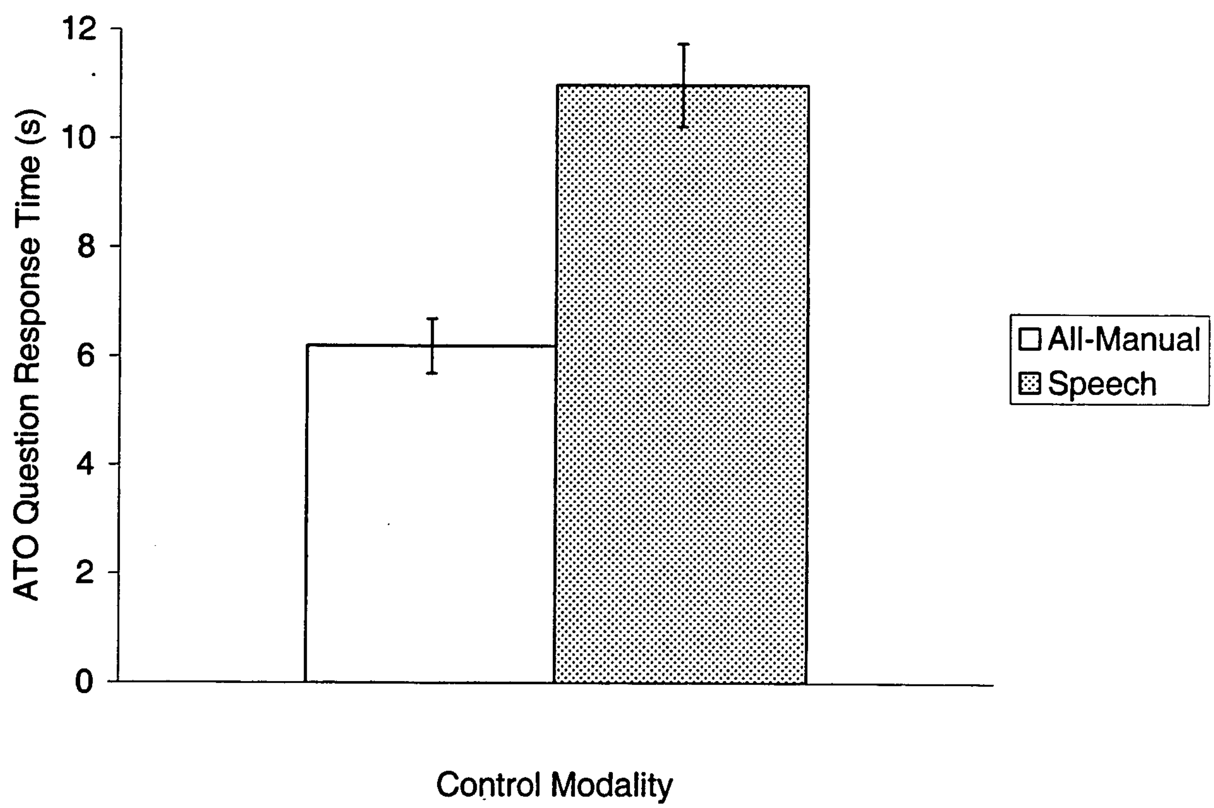

Figure 7. The mean times and standard errors to answer the Senior Director's probe questions about the Air Tasking Order (ATO) as a function of control modality.

Communications Behavior

The number of communications events and the total duration across all events within a trial were collected. The events were categorized according to the identity of the speaker (i.e., the participant, SD, and Strike) and in the case of the participant the events were subcategorized according to the communications channel used (i.e., speaking to the computer via the speech recognizer or to the white team via the normal intercom).

Participant Communications. The total number of participants' white team communications events were analyzed with a 2 (Control Modality) × 3 (Block) repeated measures ANOVA, which detected a significant main effect of Control Modality, $F(1, 11) = 6.325, p = .029$). The participants spoke to white team members a total of just of 3 more times in the speech trials (48.6 communications events) than in the manual trials (45.5). The analysis of the total speaking duration also found that the participants spent significantly more time speaking to the white team in speech trials (mean = 221.5 s) than in the manual trials (208.3 s), $F(1, 11) = 6.422, p = .028$.

The participants, of course, did not speak to the computer during the manual control modality trials. The participants' communication events with the computer during speech control modality trials were analyzed with two one-way (Block) ANOVAs. The mean number of communications events involving the participant speaking to the computer during speech control modality trials was 48.8 events for an average total duration during the trial of 148.3 s. There was no significant effect of block in either analysis.

SD Communications. The SD's communication events with the participant were analyzed with two Control Modality × Block ANOVAs. The mean number of communications events involving SD speaking to the participant over the intercom was 19.2 events for an average total duration during the trial of 145.7 s. There was no significant effect of control or block (or their interaction) in either analysis.

Strike Communications. The Strike's communication events with the participant were analyzed with two Control Modality × Block ANOVAs. The mean number of communications events involving Strike speaking to the participant over the intercom was 22.6 events for an average total duration during the trial of 40.6 s. There was no significant effect of control or block (or their interaction) in either analysis.

Intercom Speech Intelligibility

The participant's sensitivity to information presented over the intercom system was assessed by the coordinate response measure (CRM) task. There was no significant main effect of control modality or bock on the percentage of correct responses to the CRM task ($p > .05$). The Control Modality × Block interaction also failed to reach statistical significance.

TDF/PAD Interface Behavior

The number of several basic interface activities was recorded on every trial of the main experimental block. The categories included pairing events (i.e., initial pairings, repairings, and pairing terminations), hooks (i.e., "selecting" a display entity to either display more information about it or in preparation for some other action, such as pairing), and eye-point changes (e.g., changing the center-point of the display or zooming). Each category was analyzed with a 2 (Control Modality) × 3 (Block) ANOVA. The mean number of each of these behaviors as a function of Control Modality are displayed in Figure 8.

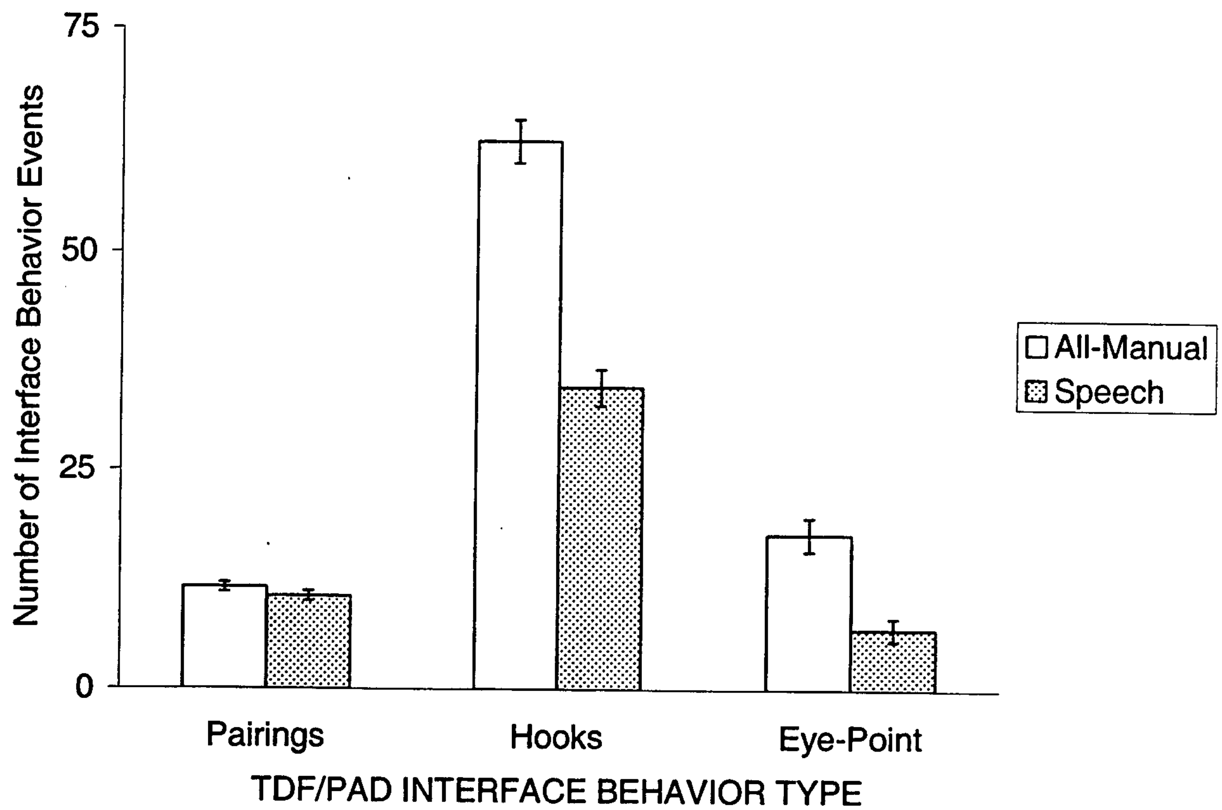

Figure 8. The mean number and standard errors of three TDF/PAD interface activity measures as a function of control modality.

Pairing Events. There was a statistically significant main effect of Control Modality on the total number of pairing events performed in each trial, $F(1, 11) = 8.609, p = .014$. More pairing events were performed in the manual control modality trials (11.6) than in the speech control modality trials (10.6). Neither the main effect of block nor the Control Modality × Block interaction reached significance.

Hooks. There was a statistically significant main effect of Control Modality on the total number of hooks performed in each trial, $F(1, 11) = 98.522, p < .001$. More hooks were performed in the manual control modality trials (62.6) than in the speech control modality trials (34.5). Neither the main effect of block nor the Control Modality × Block interaction reached significance.

Eye-point Changes. There was a statistically significant main effect of control modality on the total number of eye-point changes performed in each trial, $F(1, 11) = 26.322, p < .001$. More eye-point changes were performed in the manual control modality trials (17.7) than in the speech control modality trials (6.9). Neither the main effect of block nor the Control Modality × Block interaction reached significance.

Preference Trial Selections

Recordings of the preference trials were reviewed and the participants' choices of input modalities for all speech-enabled tasks were recorded. A task might be performed completely manually, completely by speech, or by a combination of manual and speech inputs. The percentages of participants using only speech or speech in combination with manual inputs on the 13 speech-enable tasks are listed in Table 1. The percentages range from a low of 10% to a high of 100%, with only three tasks below the 50% level.

Table 1. The Percentage of Participants Using Speech to Accomplish Speech-Enabled Tasks During the Preference Trials

TASK	%
Set-up Phase, Open ATO	50
Set-up Phase, Mark Controller	100
Set-up Phase, Sort ATO	64
Set-up Phase, Set Bulls-eye	91
Ingress Phase, Hook Aircraft (Tag)	100
Ingress Phase, Check-in Aircraft	67
Ingress Phase, Open ATO	18
Ingress Phase, Hook Aircraft (Pair)	82
Retargeting Phase, Repairing	100
Switch Bulls-eye Probe Task	83
Range & Bearing Probe Task	58
ATO Question Probe Task	42
Threat Call Hooks	10

Debriefing Questionnaire Responses

All of the participants' ratings and qualitative responses from the Post-Experimental Speech Interface Survey Form are recorded in Appendix A.

Quantitative Performance Ratings. On the four Performance Rating questions, the overall pattern of the result was highly favorable to the speech inputs. The average responses for the four questions are listed in Table 2.

Table 2. The Mean Ratings and Standard Errors (s.e.) for the Four Performance Rating Questions on the Post-Experimental Speech Interface Survey Form

Performance Rating Question	*Mean (s.e.)*
"Did the speech interface help or hinder...?"	3.5 (0.31)
"Was speech command vocabulary appropriate...?"	2.3 (0.53)
"How much confidence did you have...?"	2.5 (0.70)
"How easy was it to use the speech interface?"	2.7 (0.58)

Quantitative Utility Ratings. On the four Utility Rating Questions, the results favored application of the speech inputs as an option to the four tasks identified. The results are summarized in Table 3 (Note: One participant did not respond to the Utility Rating question regarding bearing and range questions). On the first three questions 80% or better of the participants rated speech "Somewhat Useful" or better. However, the most favorable reaction was to the final question concerning the application of speech for pairing tracks. For this task over 90% of the participants rated the application of speech "Very Useful" or better.

Table 3. The Number of Participants that Selected each Response Category for the Four Utility Rating Questions on the Post-Experimental Speech Interface Survey Form

Utility Rating Question	*Waste of Effort*	*Minor Utility*	*Somewhat Useful*	*Very Useful*	*Extremely Useful*
Interacting with ATO	0	1	3	7	1
Interacting with TDF/PAD situation display	0	0	9	3	0
Responding to bearing & range query	0	2	2	6	1
Pairing friendly tracks against targets	0	1	0	3	8

Qualitative Survey Responses. The range of responses recorded in Appendix A defies simple summarization other than for a few easily identifiable trends. For example, most evaluative comments were very favorable to the application of speech to the air battle management task. This is very evident in the responses to Question 2 (Overall Opinion of Speech Controls) where all but a few participants were very favorable in their comments. The "Good Uses" for speech identified in response to Question 3 showed evidence that the operators had reached some consensus regarding the application of speech controls. The use of speech in hooking and pairing tracks and working with the ATO were common themes in the written comments. Written comments on "Bad Uses" of speech were less frequent and did not display as much agreement across participants.

DISCUSSION

The results of this research strongly support the application of speech recognition to the TDF/PAD interface for use in AWACS applications. The use of speech as a control modality allowed operators to perform some tasks, such as system set-up or the aircraft-target repairings, more quickly than they could using the baseline manual controls. There was also evidence that the availability of speech controls allowed for more efficient time-sharing performance. This was suggested by the faster nine-line transmission times observed in the speech-enabled trials. Speech recognition was not used in the acquiring of the new nine-line from SD, writing it down, and transmitting it to Strike, but nevertheless these tasks were performed more quickly when speech controls were enabled.

The operators that participated in the study also strongly favored the use of speech controls. This was demonstrated not only by their subjective workload ratings and questionnaire responses, but most directly by their behavior in the preference trials. Given a choice of using or not using speech for the speech enabled tasks, the operators overwhelmingly chose to do so.

The quantified TDF/PAD interface behaviors (i.e., pairing events, hooks, and eye-point changes) demonstrated a considerable, and statistically significant reduction, of manual control workload. Given that some control event had shifted to speech, the reduction in manual workload, though important, is not surprising. However, it is interesting that using the speech controls for an average of almost 150 seconds during trials did not lessen the operators' speaking to SD and Strike. In fact, there was a small (but statistically significant) increase in the number of events and amount of time that the participant spoke to SD or Strike.

Overall, the results support the observation that humans are excellent using their response systems to satisfy multiple simultaneous constraints (e.g., McClelland, Rumelhart, & Hinton, 1986). Just as the typical person naturally changes his or her reaching and grasping movements to match the item being reached for (e.g., pencil or coffee mug), the operators in this experiment naturally made use of the speech and manual controls to optimize task performance, especially in the preference trials.

Although the results were strongly positive regarding the uses of speech in the AWACS HMI, there were some cautionary results as well. The manual control benefit observed in the ATO question probe task clearly shows that speech input is not a panacea for all tasks. Also, the participants clearly identified on the debriefing questionnaire (see Appendix A) several tasks that they would not consider to be candidates for speech controls.

In order to take advantage of the potential benefits of adding speech recognition controls to the TDF/PAD interface, it is necessary to conduct a more thorough analysis of the best vocabulary and grammar of speech commands to support the ABM mission. The current study only used a small subset of easily-implemented commands, but could be useful as a starting point. Of course, the participants comments on good and bad uses of speech (Appendix A) provides a good starting point for expanding the vocabulary and grammar.

It is also important to note that using speech recognition as an alternative control modality is just the beginning of the potential utility of the technology in the ABM environment. Given the recognition capabilities of current speech technology, it is reasonable to speculate that real-time transcription of verbal communications both within the AWACS and between the AWACS and outside entities is a possibility. Not only would this provide a valuable database of real-time mission adaptation in the operational environment, but the value of automatic recording of verbalizations in team research would be very great.

REFERENCES

Air Ministry Account (1941). *The Battle of Britain*. London: His Majesty's Stationery Office.

Armistead, E. L. (2002). *AWACS and HAWKEYES*. St Paul, MN: MBI Publishing.

Bolia, R.S. (2003). Effects of spatial intercoms and active noise reduction headsets on speech intelligibility in an AWACS environment. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 100-103). Santa Monica, CA: Human Factors and Ergonomics Society.

Bolia, R.S., Nelson, W.T., Ericson, M.A., & Simpson, B.D. (2000). A speech corpus for multi-talker communications research. *Journal of the Acoustical Society of America, 107*, 1065-1066.

Bortolussi, M.R., & Vidulich, M.A. (1989). The benefits and costs of automation in advanced helicopters: An empirical study. In *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 594-599). Columbus: The Ohio State University Aviation Psychology Laboratory.

Bortolussi, M.R., & Vidulich, M.A. (1991). The effects of speech controls on performance in advanced helicopters in a double stimulation paradigm. In *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 216-220). Columbus: The Ohio State University Aviation Psychology Laboratory.

Boyne, W.J. (2003). *Operation Iraqi Freedom: What went right, what went wrong, and why*. New York: Forge.

Conn, E. S. (2003). *JAVA-BASED performance Oriented visualization System* (Tech. Rep. No. AFRL-IF-RS-TR-2003-87). Wright-Patterson Air Force Base, OH: Air Force Research Laboratory.

Dictionary of Military and Associated Terms (JP 1-02). Department of Defense. <<www.dtic.mil/doctrine/jel/new_pubs/jp1_02.pdf>>. 8 Dec 2003.

Draper, M., Calhoun, G., Ruff, H., Williamson, D., & Barry, T. (2003). Manual versus speech input for unmanned aerial vehicle control station operations. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 109-113). Santa Monica, CA: Human Factors and Ergonomics Society.

Fahey, R. P., Rowe, A. L., Dunlap, K. L., deBoom, D. O. (2001). *Synthetic task design: cognitive task analysis of AWACS weapons director teams* (Tech. Rep. No. AFRL-HE-AZ-TR-2000-0159). Wright-Patterson Air Force Base, OH: Air Force Research Laboratory.

Hirst, M. (1983). *Airborne early warning: Design, development and operations*. London, Osprey.

Isby, D.C. (1997). *Jane's fighter combat in the jet age*. London: Harper Collins.

Judge, P., & Browder, S. (1998, February 23). Let's talk: Speech technology is the next big thing in computing. *Business Week,* 60-80.

Klinger, D. W., Andriole, S. J., Militello, L. G., Adelman, L., & Klein, G. (1993). *Designing for performance: a cognitive systems engineering approach to modifying an AWACS human computer interface* (Tech. Rep. No. AL/CF-TR-1993-0093). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

McClelland, J.L. Rumelhart, D.E., & Hinton, G.E. (1986). The appeal of parallel distributed processing. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel* *distributed processing: Explorations in the microstructure of cognition – Volume 1: Foundations* (pp. 3-44). Cambridge, MA: MIT Press.

McMillan, G., Eggleston, R., & Anderson, T. (1997). Nonconventional Controls. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 729–771). New York: John Wiley & Sons, Inc.

Nelson, T., Vidulich, M., Bolia, R., McLaughlin, A., Guilliams, N., & Donnelly, B. (2003). Designing speech interfaces for command and control applications. *Proceedings of the Sixth Annual International Australian Aviation Psychology Symposium.* Victoria, Australia: Australian Psychology Aviation Association.

Proctor, P. (2000, February 21). Boeing designs flexible voice controlled JSF panel. *Aviation Week & Space Technology, 152*(8), 107-110.

Roush, W. (2003, June). Computers That Speak Your Language. *Technology Review*, 32-39.

Salisbury, M., & Chilcote, J. (1989, Oct./Nov.). Investigating voice I/O for the Airborne Warning and Control System (AWACS). *Speech Technology, 5*, 50-55.

Sweeny, M. M. (2002). *An introduction to command and control.* Unpublished master's thesis, Naval Postgraduate School, Monterey, CA.

Vidulich, M.A. (1988). Speech responses and dual-task performance: Better time-sharing or asymmetric transfer? *Human Factors, 30*, 517-529.

Vidulich, M.A., & Bortolussi, M.R. (1988a). A dissociation of objective and subjective workload measures in assessing the impact of speech controls in advanced helicopters. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 1471-1475). Santa Monica, CA: Human Factors Society.

Vidulich, M.A., & Bortolussi, M.R. (1988b). Speech recognition in advanced rotorcraft: Using speech controls to reduce manual control overload. In *Proceedings of the American Helicopter Society National Specialists Meeting – Automation Applications in Rotorcraft.* Atlanta: GA: American Helicopter Society.

Vidulich, M.A., Bolia, R.S., Nelson, W.T. (in press). Technology, organization, and collaborative situation awareness in air battle management: Historical and theoretical perspectives. To appear in, S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory, measures and application.* London: Ashgate.

Vidulich, M.A., & Wickens, C.D. (1981). *Time-sharing manual control and memory search: The joint effects of input and output modality competition, priorities and control order* (Tech. Rep. No. EPL-81-4/ONR-81-4). Champaign, IL: Engineering-Psychology Research Laboratory.

Wickens, C.D. (1980). The structure of attentional resources. In R.S. Nickerson (Ed.), *Attention and Performance VIII* (pp. 239-257). Hillsdale, NJ: Erlbaum.

Wickens, C.D., Sandry, D.L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors, 25*, 227-248.

Wickens, C.D., Vidulich, M., Sandry, D., & Schiflett, S. (1981). Factors influencing the performance advantage of speech technology. In *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 705-709). Santa Monica, CA: Human Factors Society.

Wickens, C.D., Vidulich, M., & Sandry-Garza, D. (1984). Principles of S-C-R compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. *Human Factors, 26*, 533-543.

Williamson, D.T. (1999). Robust speech recognition interface to the electronic crewmember: Progress and challenges. In J. Reising, R. Taylor, & R. Onken (Eds.), *The human electronic crew: The right stuff?* (Tech. Rep. No. AFRL-HE-WP-TR-1999-0235)(pp. 115-121). Wright-Patterson Air Force Base, OH: Human Effectiveness Directorate.

Williamson, D. T., & Barry, T. P. (2000). The design and Evaluation of speech interface for generation of air tasking orders. *Proceedings of the IEA 2000/HFES 2000 Congress*, vol. 3, 750 -753. Santa Monica, CA: Human Factors and Ergonomics Society.

Williamson, D.T., Barry, T.P., & Draper, M.H. (in press). Commercial speech recognition technology in the military domain: Results of two recent research efforts. To appear in, *Proceedings of the 2004 Applied Voice Input/Output Society Conference*.

APPENDIX A:
POST-EXPERIMENTAL SPEECH INTERFACE SURVEY FORM RESPONSES

QUESTION 1: Overall Instructions – Please complete the following question by circling the appropriate rating.

QUESTION (1A) - Did the speech interface help or hinder your performance in this experiment, and by how much?

-5	-4	-3	-2	-1	0	1	2	3	4	5
much more difficult					no effect					much easier

In what way(s) did it help or hinder?

Participant 1 (Rating = 4)

The speech function almost eliminated the need to type. This allowed quicker pairing of assets to targets, location of fighters.

Participant 2 (Rating = 4)

It helped in being able to hook faster, commit and setting up the ATO and the data base while players check in.

Participant 3 (Rating = 2)

Helped hooking. So-so w/ATO. ATO vocabulary library may need more refinement than commands.

Participant 4 (Rating = 4)

Help check-in w/out leaving data screen, helpful tracking.

Participant 5 (Rating = 2)

In some aspects it make things a lot easier.

Participant 6 (Rating = 3)

It was a good way to set up ATO by controller, also a good way to find aircraft by call sign.

Participant 7 (Rating = 3)

Easier to pair to targets. Exact phrase and no real indication if it didn't accept the input. For callsigns used full name instead of first and last letter; would be quicker in the CS search box. Possibly a scroll left, right, up, down to help move around @ smaller zooms. Never zoomed in/out due to too much problem moving around the scope. Didn't like auto center for CS search; could just highlight if on screen.

Participant 8 (Rating = 4)

Made it possible to do multiple tasks at the same time. Also locating items become much greater.

Participant 9 (Rating = 4)

Faster.

Participant 10 (Rating = 5)

I was able to hook and tag aircraft coming in with a package more effectively because the computer found the aircraft and hooked it for me.

Participant 11 (Rating = 2)

It isn't really formatted in a practical way – The SD and strikes on the same freq. vs. the computer on another – real world it would be much more difficult.

Participant 12 (Rating = 5)

After using the correct verbiage it was easier and faster to use than manual.

QUESTION (1B) – Was the speech command vocabulary appropriate for the tasks that you were asked to perform?

-5	-4	-3	-2	-1	0	1	2	3	4	5
completely inappropriate					neither					perfect

Participant 1 (Rating = 2)

The vocabulary was adequate, but some of the speech commands were too time consuming (sort call sign priority 2 ascending)

Participant 2 (Rating = 2)

No written comments.

Participant 3 (Rating = 1)

Needs refinement. Needs syntax memorization. Neither good nor bad

Participant 4 (Rating = 4)

No written comments.

Participant 5 (Rating = 2)

No written comments.

Participant 6 (Rating = 2)

No written comments.

Participant 7 (Rating = 2)

No written comments.

Participant 8 (Rating = 4)

No written comments.

Participant 9 (Rating = 2)

New language to memorize

Participant 10 (Rating = 4)

No written comments.

Participant 11 (Rating = -2)

No written comments.

Participant 12 (Rating = 5)

No written comments.

<u>QUESTION (1C) – How much confidence did you have in using the speech interface?</u>

-5	-4	-3	-2	-1	0	1	2	3	4	5
low confidence					neutral					high confidence

Participant 1 (Rating = 4)

Once I had a chance to practice, my confidence greatly improved.

Participant 2 (Rating = 3)

No written comments.

Participant 3 (Rating = 3)

No written comments.

Participant 4 (Rating = 4)

No written comments.

Participant 5 (Rating = -1)

No written comments.

Participant 6 (Rating = 4)

No written comments.

Participant 7 (Rating = 3)

No written comments.

Participant 8 (Rating = 3)

No written comments.

Participant 9 (Rating = -2)

Made mistakes/not reliable for checklists.

Participant 10 (Rating = 5)

No written comments.

Participant 11 (Rating = -1)

No written comments.

Participant 12 (Rating = 5)

No written comments.

QUESTION (1D) – How easy was it to use the speech interface?

-5	-4	-3	-2	-1	0	1	2	3	4	5
not easy					neutral					very easy

Participant 1 (Rating = 5)

Given a chance to practice, the speech interface was very easy.

Participant 2 (Rating = 4)

No written comments.

Participant 3 (Rating = 2)

No written comments.

Participant 4 (Rating = 4)

No written comments.

Participant 5 (Rating = -1)

No written comments.

Participant 6 (Rating = 1)

No written comments.

Participant 7 (Rating = 3)

No written comments.

Participant 8 (Rating = 4)

No written comments.

Participant 9 (Rating = 3)

No written comments.

Participant 10 (Rating = 4)

No written comments.

Participant 11 (Rating = -1)

No written comments.

Participant 12 (Rating = 4)

No written comments.

QUESTION (2) - Overall Opinion of Speech Controls – Considering the speech controls that you used today, what is your overall opinion regarding the possibility that a completed and fully debugged version of speech controls would be a desirable feature for future AWACS interfaces?

Participant 1

Absolutely. The speech controls would be an excellent addition to future AWACS interfaces. Given the fact that future command and control will be even more focused on time sensitive tasks, speech controls will allow future AWOs the ability to handle more complex and a greater number of tasks in the same amount of time.

Participant 2

Very desirable, it would make a lot of work load go away during critical parts of flight/controlling. It's much easier to talk commands rather than having to do switch actions. I find it faster to talk then having to mess around w/the computer.

Participant 3

There are some commands that are very good vocally. Some are not. Provision should be made for manual actions when voice in not expedient. There's no "undo" for mistakes.

Participant 4

I thought it worked great. It shows surprising capabilities. I think it would be very useful on the jet.

Participant 5

A lot of changes would have to be made to the AWACS system. Tracking is a huge deficiency on the AWACS. Speech interface would not work if tracking was not perfect. Also, during a task saturated environment, it would be impossible to use. It is also not designed to be versatile

Participant 6

I feel that speech controls would be very helpful in time critical situations. However, whenever it picks up the wrong thing it is very frustrating

Participant 7

See question 1. Think it would be great for organizing admin items (i.e. sto – which isn't on jet currently), bringing up airspace LOA items and replacing/make easier free text messaging and sending items via link. Imagine a station handover w/ an excel type spreadsheet listing in AOR players/status sent via link to the incoming E-3. Does increase noise levels on jet and reduces "NET 4" SA building. One advantage to typing is ability to type info while talking on radio's. Would use talking time to do a commit, while now you could be typing msg.

Participant 8

Combined with manual controls I think it would greatly enhance the job. Tailoring it to multiple voices and intonations will probably be difficult, but if you can I see it as a very useful tool.

Participant 9

I don't think it is foolproof yet. Had to speak at a slow pace and only took sometimes.

Participant 10

I would really like having this option because it cuts time needed to do switch actions on the jet.

Participant 11

It would be necessary to have it fully debugged and to have maximum AWO input as to what was practical for speech and need requirements. I think it should be an option, not necessarily a must in case it doesn't work for some.

Participant 12

I believe once all and any major limitation is removed from the interface, the voice command program would be very effective on the AWACS. I believe it would save valuable time for the AWO.

QUESTION (3) - Good and Bad Uses of Speech Controls – Based on your experiences today and your background in AWACS, please identify any applications of speech recognition that you think would be especially good or bad in an AWACS environment:

GOOD USES FOR SPEECH　　　　BAD USES FOR SPEECH

Participant 1

Good Uses:

ATO check-in; 1-Hooking Fighters; Committing fighters to targets

Bad Uses:

Bringing up (show) bull's-eye (no quicker, actually slower than w/ a mouse)

Participant 2

Good Uses:

ATO setup; Committing; Finding players; Set-up data base; Set-up of Scope; Updating players

Bad Uses:

Adding/dropping players; B/E and Threat Calls; ATO sorting

Participant 3

Good Uses:

ATO Search; Hook into a cloud of symbology; Retargeting!! (assumes tgt name in d-base AND correct); Expand out (zoom)

Bad Uses:

Initial Pairing; Closing windows; Expand in (Zoom)

Participant 4

Good Uses:

Commits; Targets; Database interaction; Quickly locating Track (ie: hook)

Bad Uses:

Zoom

Participant 5

Good Uses:

Committing Symbology; Hooking Tracks (very good!!); Hide and show B/E

Bad Uses:

Hide and show B/E

Participant 6

Good Uses:

Finding Symbology; Pairing A/C or committing them; Setting up B/E

Bad Uses:

Expanding and setting view

Participant 7

Good Uses:

Pairing

Bad Uses:

Utilizing computer while talking on radio

Participant 8

Good Uses:

Faster locating items; multi-tasking

Bad Uses:

relying on it too much; different accents

Participant 9

Good Uses:

Set-up

Bad Uses:

Emergencies; Checklists

Participant 10

Good Uses:

Hook option; open/sort ATO; commit us

Bad Uses:

No written comments.

Participant 11

Good Uses:

Hooking track w/o having to search for it; Having it commit to tgt. w/o searching; Showing hiding BE/ coordinates

Bad Uses:

Can't use "show" ATO; Can't close window i.e., ATO; Accents???; Frustration i.e., faster/slower talk

Participant 12

Good Uses:

Opening lists; Hooking correct participants; pairing

Bad Uses:

if limitations on speech recognition removed, or fixed or minor, I don't see any bad uses for it

QUESTION 4: Utility Ratings Overall Instructions – Please rate the utility of the following possible uses of speech recognition. Just place a mark in the box above the best description. Note that not all of these options were demonstrated today, but they do represent possible uses of speech recognition technology.

QUESTION (4A) – Speech Recognition for interacting with the ATO.

Waste of Effort	Minor Utility	Somewhat Useful	Very Useful	Extremely Useful
I would never use speech for this.	Speech would help on rare occasions.	Speech or manual would be equally used.	Speech would be generally preferred.	Probably would not use manual at all.

Participant 1 (Rating = 4)

Check-in greatly improved, sorting and responding to ATO queries were hindered by using speech

Participant 2 (Rating = 4)

No written comments.

Participant 3 (Rating = 3)

No written comments.

Participant 4 (Rating = 3)

No written comments.

Participant 5 (Rating = 2)

No written comments.

Participant 6 (Rating = 4)

No written comments.

Participant 7 (Rating = 3)

No written comments.

Participant 8 (Rating = 4)

No written comments.

Participant 9 (Rating = 5)

No written comments.

Participant 10 (Rating = 4)

No written comments.

Participant 11 (Rating = 4)

No written comments.

Participant 12 (Rating = 4)

No written comments.

QUESTION (4B) – Speech Recognition for interacting with the TDF situation display.

Waste of Effort	Minor Utility	Somewhat Useful	Very Useful	Extremely Useful
I would never use speech for this.	Speech would help on rare occasions.	Speech or manual would be equally used.	Speech would be generally preferred.	Probably would not use manual at all.

Participant 1 (Rating = 4)

No written comments.

Participant 2 (Rating = 3)

No written comments.

Participant 3 (Rating = 3.5)

No written comments.

Participant 4 (Rating = 4)

No written comments.

Participant 5 (Rating = 3)

No written comments.

Participant 6 (Rating = 3)

No written comments.

Participant 7 (Rating = 3)

No written comments.

Participant 8 (Rating = 4)

No written comments.

Participant 9 (Rating = 3)

No written comments.

Participant 10 (Rating = 3)

No written comments.

Participant 11 (Rating = 3)

No written comments.

Participant 12 (Rating = 3)

No written comments.

QUESTION (4C) – Speech Recognition for responding to the bearing and range query.

Waste of Effort	Minor Utility	Somewhat Useful	Very Useful	Extremely Useful
I would never use speech for this.	Speech would help on rare occasions.	Speech or manual would be equally used.	Speech would be generally preferred.	Probably would not use manual at all.

Participant 1 (Rating = 5)

No written comments.

Participant 2 (Rating = 4)

No written comments.

Participant 3 (Rating = No response*)*

No written comments.

Participant 4 (Rating = 4)

No written comments.

Participant 5 (Rating = 2)

No written comments.

Participant 6 (Rating = 4)

No written comments.

Participant 7 (Rating = 4)

No written comments.

Participant 8 (Rating = 3)

No written comments.

Participant 9 (Rating = 4)

No written comments.

Participant 10 (Rating = 3)

No written comments.

Participant 11 (Rating = 2)

No written comments.

Participant 12 (Rating = 4)

No written comments.

QUESTION (4D) – Speech Recognition for pairing friendly tracks against targets.

Waste of Effort	Minor Utility	Somewhat Useful	Very Useful	Extremely Useful
I would never use speech for this.	Speech would help on rare occasions.	Speech or manual would be equally used.	Speech would be generally preferred.	Probably would not use manual at all.

Participant 1 (Rating = 5)

No written comments.

Participant 2 (Rating = 5)

No written comments.

Participant 3 (Rating = 4)

No written comments.

Participant 4 (Rating = 4)

No written comments.

Participant 5 (Rating = 5)

No written comments.

Participant 6 (Rating = 5)

No written comments.

Participant 7 (Rating = 4)

No written comments.

Participant 8 (Rating = 5)

No written comments.

Participant 9 (Rating = 2)

No written comments.

Participant 10 (Rating = 5)

No written comments.

Participant 11 (Rating = 5)

No written comments.

Participant 12 (Rating = 5)

No written comments.